Listy Biometryczne – Biometrical Letters
Vol. 30 (1993), No. 2, 109-118

Statistical inference in two-way layouts with minimal number of observations

A. Michalski and J. Srzednicka

Department of Mathematics, Agricultural University of Wrocław

Summary

In the paper the mixed linear model coresponding to two-way layouts with the minimal number of observations is considered. Theoretical considerations as well as simulations presented in the paper prove that in this model statistical inference on variance components is "uncertain" even for the layouts that assure the existance of the uniformly best unbiased estimators and of the most powerful tests. Some suggestions are made how to improve the "efficiency" of the layouts by simple manner.

1. Two-way cross classification mixed model

Let us consider an experiment in which n experimental units are classified by two factors having r and c levels, respectively, and classification is done according to the $r{\times}c$ matrix $\mathbf{N}$ with entries $n_{ij} \geq 0$. Here n_{ij} is the number of observations in the $(i,\ j)$-th cell of units treated by the i-th level of the first (row) factor and the j-th level of the second (column) one. Let y_{ijk} be the observation taken on the k-th unit of the (i, j)-th cell, k=1,...,n_{ij}. The basic assumption is one of the additivity of the effects of the factors levels which can be written for i=1,2,...,r, j=1,2,...,c and k=1,2,...,n_{ij} , as

$$y_{ijk} = \tau_i + \gamma_j + e_{ijk}\,. \tag{1.1}$$

Here τ_i is the effect of the i-th level of the first factor, γ_j is the effect of the j-th level of the second factor, while e_{ijk} are normally distributed random errors. The above model may be represented in the following matrix form

---

Key words: two-way layouts, variance components, invariant tests for variance components

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\gamma} + \mathbf{X}_2\boldsymbol{\tau} + \mathbf{e}, \tag{1.2}$$

where $\mathbf{y}$ is an $n\times1$-vector of observations, $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_c)'$, $\boldsymbol{\tau} = (\tau_1,\ldots,\tau_r)'$, $\mathbf{e}$ is the vector of errors and $\mathbf{X}_1$, $\mathbf{X}_2$ are $n\times c$ and $n\times r$ known matrices of full ranks r and c, respectively, which elements are 0 or 1 depending on the ordering of the components of $\mathbf{y}$. Anywhere $\mathbf{X}_2'\mathbf{X}_1 = \mathbf{N}$, $\mathbf{X}_1\mathbf{1}_c = \mathbf{X}_2\mathbf{1}_r = \mathbf{1}_n$, while $\mathbf{D}_c = \mathbf{X}_1'\mathbf{X}_1 = \mathrm{diag}\{n_{.j}\}$ and $\mathbf{D}_r = \mathbf{X}_2'\mathbf{X}_2 = \mathrm{diag}\{n_{i.}\}$ are diagonal matrices with diagonal elements $n_{i.}$ and $n_{.j}$, respectively, where $n_{i.} = \Sigma_j n_{ij}$ and $n_{.j} = \Sigma_i n_{ij}$. We will assume that $n_{i.}$ and $n_{.j}$ are all positive so that both $\mathbf{D}_r$ and $\mathbf{D}_c$ are positive definite. In many applications of two-way classification model effects are considered to be random. The assumption that τ is normally distributed with $E(\boldsymbol{\tau}) = \mathbf{0}$, $E(\boldsymbol{\tau\tau}') = \sigma_\tau^2\mathbf{I}_r$, $E(\boldsymbol{\tau}\mathbf{e}') = \mathbf{0}$ leads to the mixed model in which

$$E(\mathbf{y}) = \mathbf{X}_1\boldsymbol{\gamma}; \quad \mathrm{Cov}(\mathbf{y}) = \sigma_\tau^2\mathbf{X}_2\mathbf{X}_2' + \sigma_e^2\mathbf{I}_n. \tag{1.3}$$

The general problem we consider in the paper is a statistical inference on the variance components σ_τ^2 and σ_e^2. The main considerations are restricted to layouts with the minimal number of observations. Theoretical considerations as well as simulations presented in the paper prove that in the mixed model (1.3) corresponding to layouts with the minimal number of observations statistical inference on variance components is "uncertain" even for the models that assure the existance of the uniformly best unbiased estimators and of the most powerful tests. Some suggestions are made how to improve the "efficiency" of the layouts by simple manner.

2. Layouts with the minimal number of observations

Let us consider an $r\times r$-matrix $\mathbf{C}_r$ defined as follows

$$\mathbf{C}_r = \mathbf{X}_2'\mathbf{M}_c\mathbf{X}_2 = \mathbf{D}_r - \mathbf{N}\mathbf{D}_c^{-1}\mathbf{N}' \tag{2.1}$$

with $\mathbf{M}_c = \mathbf{I}_n - \mathbf{X}_1\mathbf{D}_c^{-1}\mathbf{X}_1'$ being the orthogonal projector on the kernel of $\mathbf{X}_1'$. Since $\mathbf{C}_r\mathbf{1}_r = \mathbf{1}_r$, $\mathrm{rank}(\mathbf{C}_r) \le r-1$. Let $\mathrm{rank}(\mathbf{C}_r) = r-q$, $1 \le q \le r-1$. If q=1 then the layout is called connected, for q>1 the layout is said to be q-disconnected.

It has been proved by Kageyama (1985) (see also Dodge, 1985) that

$$n \ge c + \mathrm{rank}(\mathbf{C}_r) = r + c - q. \tag{2.2}$$

Thus $n = r + c - q$ is the minimal number of observations for the two-way layout with $\mathrm{rank}(\mathbf{C}_r) = r-q$. For construction of such layouts see Kageyama (1985). All of them are by definition binary, i.e. their incidence matrices have elements 0 and 1 only.

An interesting class of connected layouts with the minimal number of experimental units has been considered by Baksalary et al. (1990). The class consists of layouts with the incidence matrices isomorphic, with respect to permutations of rows and/or columns, to

$$\mathbf{N} = \begin{bmatrix} \mathbf{I}_{r-1} \\ ---- \\ \mathbf{1}'_{r-1} \end{bmatrix}. \tag{2.3}$$

It follows from Theorem 3 in Baksalary et al. (1990) that when $r \geq 3$ and $\mathbf{N}$ is given by (2.3) then under corresponding mixed model (1.3) there exists the uniformly best invariant quadratic and unbiased estimator (UBIQUE) for every function $f_1\sigma_\tau^2 + f_2\sigma_e^2$. Following Gnot et.al. (1992, lemma 2.2) the above model also assure the existance of the uniformly most powerful invariant test (UMPIT) for testing

$$H:\ \sigma_\tau^2 = 0 \quad \text{vs} \quad K:\ \sigma_\tau^2 > 0\,. \tag{2.4}$$

Now we present some details concerning explicit forms of the UBIQUE's and the UMPIT.

3. Estimation and testing for variance components

3.1. Estimation

The problem of estimation of $f_1\sigma_\tau^2 + f_2\sigma_e^2$ in the model (1.3) is invariant under the group $\mathcal{G}_1$ of translations $g_\gamma(\mathbf{Y}) = \mathbf{y} + \mathbf{X}_1\gamma$, $\gamma \in R^c$. Following Olsen, Seely and Birkes (1976) (see also LaMotte, 1976) a maximal invariant statistic with respect to $\mathcal{G}_1$ is $\mathbf{t} = \mathbf{By}$, where $\mathbf{B}$ is an $(n-c)\times n$ matrix defined as follows

$$\mathbf{BB}' = \mathbf{I}_{n-c}\,,\quad \mathbf{B}'\mathbf{B} = \mathbf{M}_c\,, \tag{3.1}$$

where $\mathbf{M}_c = \mathbf{I}_n - \mathbf{X}_1\mathbf{D}_c^{-1}\mathbf{X}_1'$. The model for $\mathbf{t}$ is given by

$$E(\mathbf{t}) = \mathbf{0}\,,\quad \mathrm{Cov}(\mathbf{t}) = \sigma_\tau^2\mathbf{W} + \sigma^2\mathbf{I}_{n-r}\,,\quad \mathbf{W} = \mathbf{BX}_2\mathbf{X}_2'\mathbf{B}'. \tag{3.2}$$

Denote by $\alpha_1 > \alpha_2 > \ldots > \alpha_h \geq 0$ the ordered sequence of different eigenvalues of $\mathbf{W}$ with the multiplicities $\nu_1,\ldots,\nu_h$. We shall assume through the paper that $h \geq 2$. It is easy to establish that:

(i) positive eigenvalues of $\mathbf{W}$ and $\mathbf{C}_r = \mathbf{X}_2'\mathbf{M}_c\,\mathbf{X}_2 = \mathbf{D}_r - \mathbf{ND}_c^{-1}\mathbf{N}'$ are the same,

(ii) $\alpha_h = 0$ iff $n > r+c-q$, and in this case $\nu_h = n-r-c+q$.

Let $\mathbf{W} = \sum_{i=1}^{h} \alpha_i \mathbf{E}_i$ be the spectral decomposition of $\mathbf{W}$ and $Z_i = \mathbf{t}'\mathbf{E}_i\mathbf{t} / \nu_i$, $i = 1,...,h$. The following lemma established by Olsen et al. (1976) gives basic statistical properities of $\mathbf{Z} = (Z_1,...,Z_h)'$.

Lemma 3.1

(i) $\nu_i Z_i / (\sigma_\tau^2 \alpha_1 + \sigma_e^2) \sim \chi^2_{\nu_i}$; $i=1,...,h$,

(ii) $\mathbf{Z} = (Z_1,...,Z_h)'$ is a minimal sufficient statistic for the family of distributions of t,

(iii) $\mathbf{Z}$ is a minimal complete statistic iff $h = 2$,

(iv) $Z_1,...,Z_h$ are statistically independent,

(v) for an arbitrary real vector $\mathbf{a} = (a_1,...,a_h)'$

$$\begin{cases} E(\mathbf{a}'\mathbf{Z}) = (\Sigma\, a_i \alpha_i)\sigma_\tau^2 + (\Sigma a_i)\sigma_e^2 , \\ \mathrm{Var}(\mathbf{a}'\mathbf{Z}) = 2\Sigma a_i^2 (\sigma_\tau^2 \alpha_i + \sigma_e^2)^2 / \nu_i . \end{cases} \tag{3.3}$$

Following Gnot and Kleffe (1983) admissible invariant quadratic estimators for $f_1\sigma_\tau^2 + f_2\sigma_e^2$ are linear combinations $\mathbf{a}'\mathbf{Z}$. In the paper we restrict our considerations to the following two admissible estimators:

$$\upsilon_1 = \sum_{i=1}^{h} Z_i (\lambda_1 \alpha_i + \lambda_2)\nu_1 / (1+\alpha_i)^2 , \tag{3.4}$$

$$\upsilon_2 = \sum_{i=1}^{h-1} Z_i \lambda_{01} \nu_i / \alpha_i + Z_h \lambda_{02} \nu_h . \tag{3.5}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$ and $\boldsymbol{\lambda}_0 = (\lambda_{01}, \lambda_{02})'$ are such that $E(\upsilon_i) = f_1\sigma_\tau^2 + f_2\sigma_e^2$, $i = 1,2$. It is worth noting that υ_1 is MINQUE for $f_1\sigma_\tau^2 + f_2\sigma_e^2$, while υ_2 (in the case $\alpha_h = 0$) is a limiting bayesian estimator with respect to the prior distributions κ_n, such that $\mathrm{Var}_{\kappa_n}(\sigma_\tau^2)$ tends to infinity (cf.Rao, Kleffe, 1988, pp. 309-316).

In the particular case of the model (1.3) corresponding to layout with the incidence matrix (2.3) the matrix $\mathbf{W}$ given in (3.2) is nonsingular and has two different positive eigenvalues $\alpha_1 = r/2$ and $\alpha_2 = 0.5$ with multiplicities $\nu_1 = 1$ and $\nu_2 = r-2$, respectively. There exist UBIQUE's for $f_1\sigma_\tau^2 + f_2\sigma_e^2$; in particular, for σ_τ^2 and σ_e^2 they are

$$\hat{\sigma}_\tau^2 = 2(Z_1 - Z_2) / (r-1) \quad \text{and} \quad \hat{\sigma}_e^2 = (rZ_2 - Z_1) / (r-1)$$

with

$$\mathrm{Var}(\hat{\sigma}_\tau^2) = 2\sigma_e^4 \, \frac{\theta^2(r^2-r-1) + 4\theta(r-1) + 4}{(r-1)(r-2)}$$

and

$$\mathrm{Var}(\hat{\sigma}_e^2) = 2\sigma_e^4 \, \frac{\theta^2 r^2 + 8\theta r + 4(r+2)}{4(r-1)(r-2)},$$

where $\theta = \sigma_\tau^2 / \sigma_e^2$.

3.2. Testing

The problem of testing hypothesis

$$\mathrm{H}: \ \sigma_\tau^2 = 0 \quad \text{vs} \quad \mathrm{K}: \ \sigma_\tau^2 > 0 \tag{3.6}$$

or equivalently

$$\mathrm{H}: \ \theta = 0 \quad \text{vs} \quad \mathrm{K}: \ \theta > 0, \quad \theta = \sigma_\tau^2 / \sigma_e^2 \tag{3.7}$$

in the model (1.3) is invariant under the group $\mathcal{G}_2$ of transformations $g_\gamma(\mathbf{y}) = c(\mathbf{y} + \mathbf{X}_1\gamma)$, $\gamma \in R^c$, $c \in \mathrm{R}_+$. A maximal invariant statistic with respect to $\mathcal{G}_2$ is $\mathbf{Z}_0 = (Z_1 / Z_h, \ldots, Z_{h-1} / Z_h)'$ while the Neyman-Pearson test for testing $\mathrm{H}: \ \theta = 0$ vs $\mathrm{K}: \ \theta = \theta_*$, $\theta_* > 0$ rejects H for sufficiently large

$$F_{NP}(\theta_*) = \frac{\sum_{i=1}^{h} \nu_i Z_i}{\sum_{i=1}^{h} \nu_i Z_i / (1 + \alpha_i \theta_*)} \tag{3.8}$$

(cf.Gnot et al., 1992). In the particular case when h=2 the Neyman-Pearson test coincides with the UMPIT which rejects H for sufficiently large $F = Z_1 / Z_2$. Let

$$F(\theta) = \frac{Z_1(1 + \theta\alpha_2)}{Z_2(1 + \theta\alpha_1)}. \tag{3.9}$$

Following lemma 3.1 for an arbitrary and fixed θ the statistic $F(\theta)$ has a central F-distribution with ν_1 and ν_2 degrees of freedom and $F(0) = F$.

Let $\beta(\theta)$ be the power function of the UMPIT at level p for testing (3.7) in the model (1.3) with h = 2. The following lemma gives a basic properties of the test.

Lemma 3.2. There exists $\lim_{\theta \to \infty} \beta(\theta) = \beta$ and

(i) $\beta = 1$ if $\alpha_h = 0$ $(n > r+c+q)$,

(ii) $\beta < 1$ if $\alpha_h > 0$ $(n = r+c-q)$.

Proof. First note that

$$\beta(\theta) = P_\theta(F > k_p) = P_\theta[F(\theta) > k_p(\theta)] ,$$

where k_p is determined by $\beta(0) = p$, while $k_p(\theta) = k_p(1 + \alpha_2\theta) / (1 + \alpha_1\theta)$. If θ tends to infinity then $k_p(\theta)$ tends to $k_p\alpha_2 / \alpha_1$ which is positive if $\alpha_2 > 0$.

4. Examples and simulations

According to Lemma 3.2 (ii) the power function $\beta(\theta)$ of the UMPIT for testing (3.7) in the model (1.3) with **N** given by (2.3) does not achieve 1 if θ tends to infinity. It is so because this layout has a minimal number of observations (n=r+c−1) and in consequence **W** is positive definite (the minimal eigenvalue of **W** is α=0.5). The table below shows $\lim_{\theta \to \infty} \beta(\theta)$ as a function of r for layouts with the incidence matrix having the structure (2.3); (the significance level p=0.05).

r	3	5	10	16	20	50	200	500	1000
$\beta(\infty)$	0.0862	0.2498	0.4867	0.6002	0.6441	0.7774	0.8892	0.9300	0.9505

The fact that the power function of the tests for testing (3.7) is essentially less than one even for far alternatives make the layouts with the minimal number of observations unacceptable for practice if mixed models are assumed. To overcome this inconvenience, as competitors let us consider q-disconnected layouts with the incidence matrix $\mathbf{N}^{(q)}$ having the form

$$\mathbf{N}^{(q)} = \left[\begin{array}{c|c} \mathbf{I}_{r-q} & \\ \text{------} & \mathbf{0} \\ \mathbf{1}'_{r-q} & \\ \text{------} & \text{------} \\ \mathbf{0} & 2\mathbf{I}_{q-1} \end{array}\right], \qquad 1 < q \le r-1. \tag{4.1}$$

Comparing (2.3) and (4.1) we see that the number of observations as well as the numbers of levels of two factors for both layouts are the same. However, in the model corresponding to layout with the incidence matrix (4.1) the matrix **W** given by (3.2) is singular and has three different eigenvalues $\alpha_1 = (r-q+1)/2$, $\alpha_2 = 0.5$ and $\alpha_3 = 0$ with the multiplicities $\nu_1 = 1$, $\nu_2 = r-q-1$ and $\nu_3 = q-1$.

As an example let us consider layouts with r=16 and with the incidence matrices **N** given by (2.3) and $\mathbf{N}^{(q)}$, q=2, 6, 11, given by (4.1). Figure 1 shows the power function $\beta(\theta)$ of the UMPIT in the model corresponding to **N** and the attainable upper bounds (AUB's) of the power functions in the models corresponding to $\mathbf{N}^{(q)}$, as functions of θ . The values of AUB's have been obtained as power functions of the Neyman-Pearson tests at θ (for details see LaMotte et al., 1988).

We can see that the behaviour of the power function of the locally best test in the model with $\mathbf{N}^{(q)}$ is quite different than in the case when UMPIT exists. The functions tend to 1 very fast if θ tends to infinity.

To compare the variance of estimators for the variance components σ_τ^2 and σ_e^2 we consider the UBIQUE's in the model corresponding to **N** and estimators υ_1 (MINQUE) and υ_2 (limiting bayesian estimator) given by (3.4) and (3.5) in the models with incidence matrices $\mathbf{N}^{(q)}$. Figures 2 and 3 show the ratios $R(\theta)$ of variances of MINQUE and UBIQUE for σ_τ^2 and σ_e^2, respectively, while Figures 4 and 5 show the similar ratios of variances of limiting estimator and UBIQUE. As we can see, in each presented case there exist such value θ_0, that for $\theta > \theta_0$ the variance of UBIQUE is greater than the variance of estimators in models where UBIQUE does not exist.

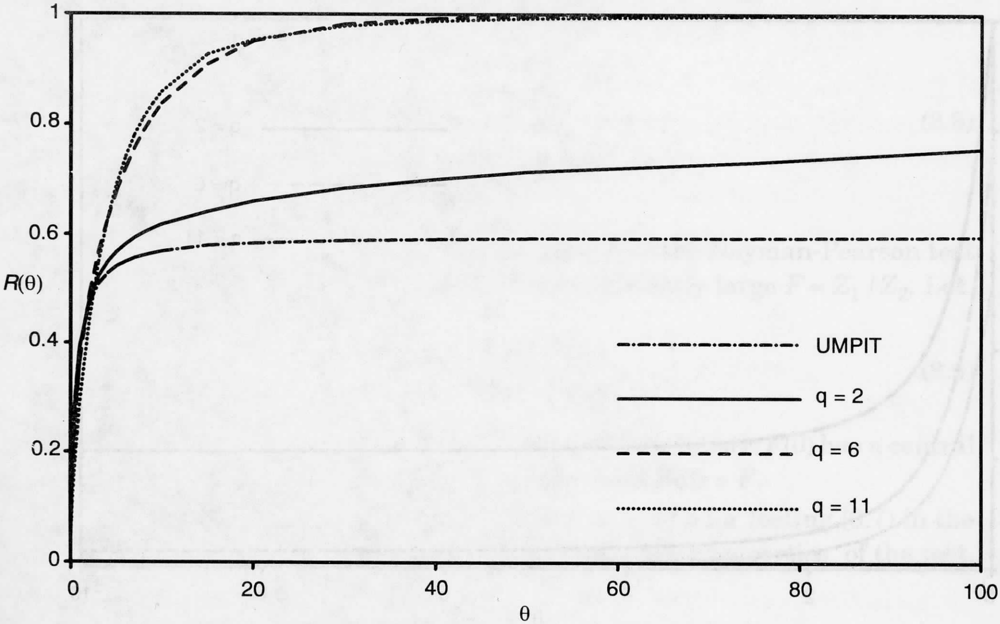

Fig. 1. Comparison of power functions

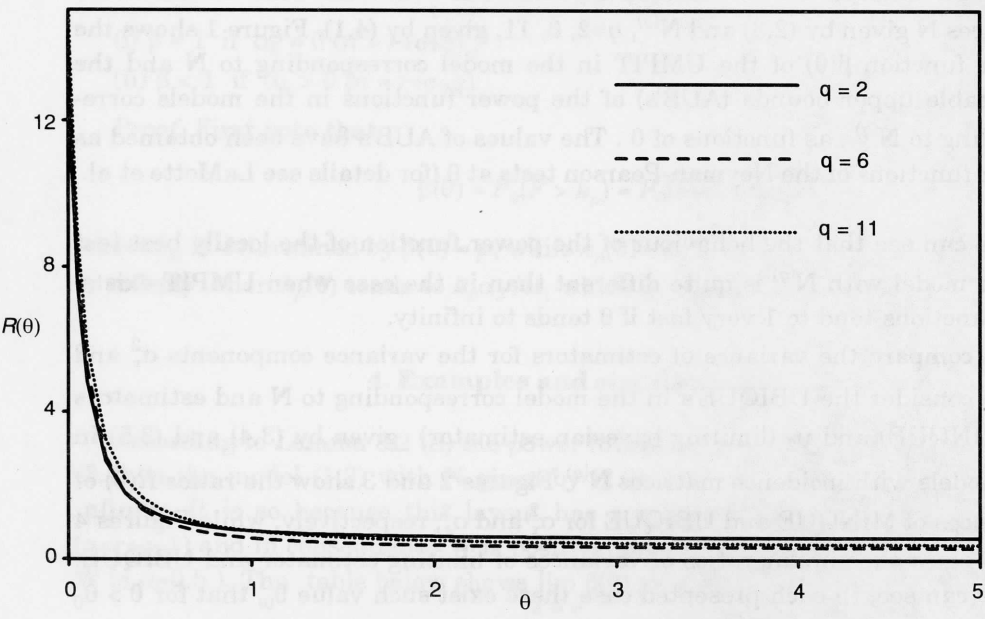

Fig. 2. Ratio of variances of MINQUE and UBIQUE for σ_τ^2

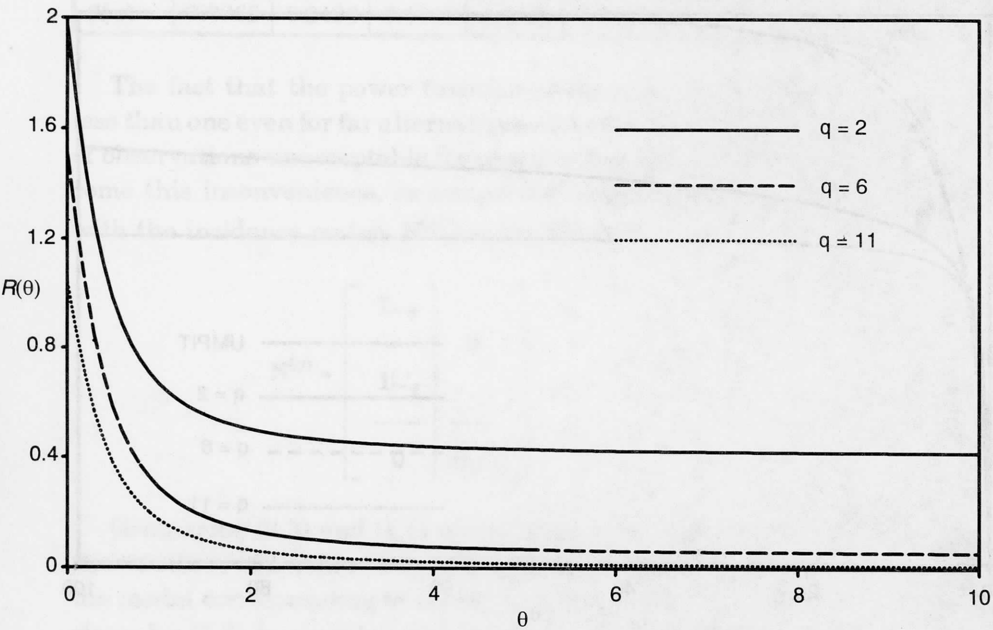

Fig. 3. Ratio of variances of MINQUE and UBIQUE for σ_e^2

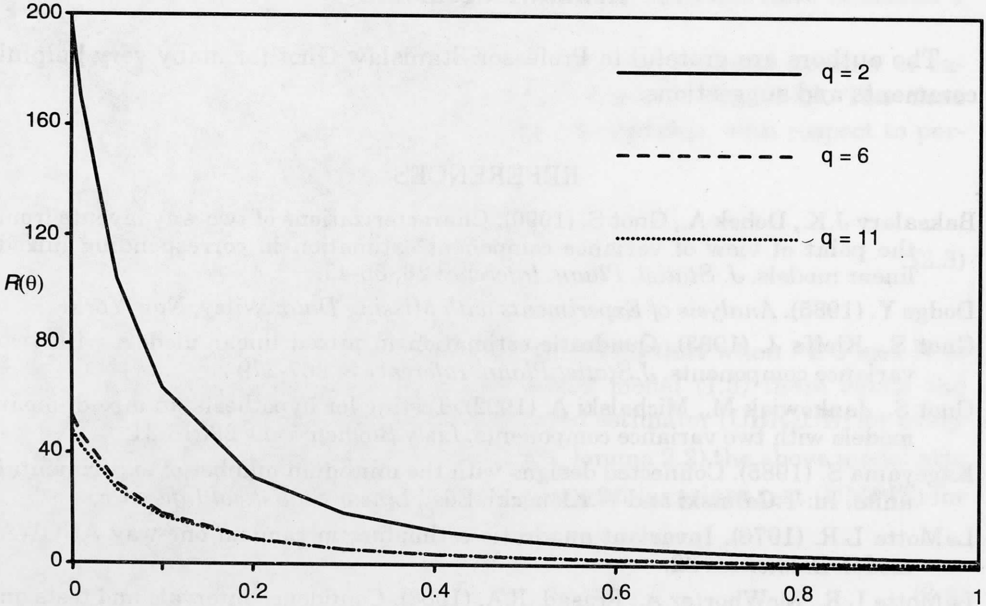

Fig. 4. Ratio of variances of limiting estimator and UBIQUE for σ_τ^2

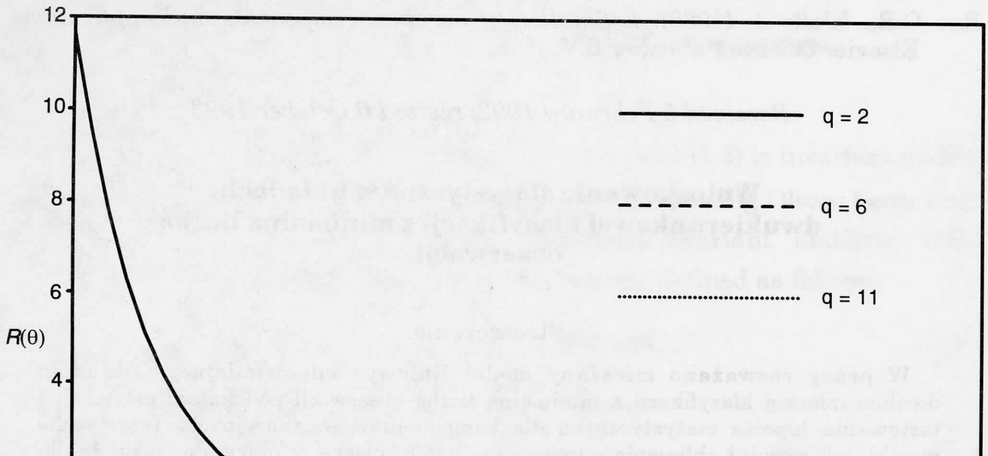

Fig. 5. Ratio of variances of limiting estimator and UBIQUE for σ_e^2

Acknowledgements

The authors are grateful to Professor Stanisław Gnot for many very helpful comments and suggestions.

REFERENCES

Baksalary J.K., Dobek A., Gnot S. (1990). Characterizations of two way layouts from the point of view of variance component estimation in corresponding mixed linear models. *J. Statist. Plann. Inference* **26**, 35-45.

Dodge Y. (1985). *Analysis of Experiments with Missing Data.* Wiley, New York.

Gnot S., Kleffe J. (1983). Quadratic estimation in mixed linear models with two variance components. *J.Statist.Plann. Inference* **8**, 267-279 .

Gnot S., Jankowiak M., Michalski A. (1992). Testing for hypothesis in mixed linear models with two variance components. *Listy Biometryczne* **29**, 13-31.

Kageyama S. (1985). Connected designs with the minimum number of experimental units. In: T.Caliński and W.Klonecki, Eds., *Linear Statistical Inference.*

LaMotte L.R. (1976). Invariant quadratic estimators in random one-way ANOWA model. *Biometrics* **32**.

LaMotte L.R., McWhorter A., Prasad, R.A. (1988). Confidence intervals and tests on the variance ratio in random models with two variance components. *Com.Stat.-Theory and Methods* **17**, 1135-1164.

Olsen A., Seely J., Birkes D. (1976). Invariant quadratic unbiased estimation for two variance components. *Ann.Statist.* **4**, 878-890.

Rao C.R., Kleffe J. (1988). *Estimation of variance components and applications.* Elsevier Science Publisher B.V.

Received 5 February 1992; revised 6 October 1993

Wnioskowanie statystyczne w układach dwukierunkowej klasyfikacji z minimalną liczbą obserwacji

Streszczenie

W pracy rozważano mieszany model liniowy odpowiadający układom dwukierunkowej klasyfikacji z minimalną liczbą obserwacji pod kątem estymacji i testowania hipotez statystycznych dla komponentów wariancyjnych. Teoretyczne wyniki, jak również obliczenia numeryczne przedstawione w pracy, dowodzą, że dla takiego modelu kryterium istnienia jednostajnie najlepszych estymatorów nieobciążonych lub jednostajnie najmocniejszego testu nie zawsze jest dobrym kryterium doboru układu. Pokazano, że niewielka modyfikacja układu doświadczalnego poprawia jego "efektywność".

Słowa kluczowe: układy dwukierunkowej klasyfikacji, komponenty wariancyjne, testy niezmiennicze dla komponentów wariancyjnych